

On the Assembly History of Dark Matter Haloes

Yun Li¹, H.J. Mo¹, Frank C. van den Bosch², W.P. Lin³ *

¹*Department of Astronomy, University of Massachusetts, MA 01003, USA*

²*Max-Planck-Institute for Astronomy, Königstuhl 17, D-69117 Heidelberg, Germany*

³*Shanghai Astronomical Observatory, 80 Nandan Rd., Shanghai 200030, China*

ABSTRACT

We study the mass assembly history (MAH) of dark matter haloes. We compare MAHs obtained using (i) merger trees constructed with the extended Press Schechter (EPS) formalism, (ii) numerical simulations, and (iii) the Lagrangian perturbation code PINOCCHIO. We show that the PINOCCHIO MAHs are in excellent agreement with those obtained using numerical simulations, while the EPS formalism predicts MAHs that occur too late. PINOCCHIO, which is much less CPU intensive than N-body simulation, can be run on a simple personal computer, and does not require any labor intensive post-simulation analysis, therefore provides a unique and powerful tool to investigate the growth history of dark matter haloes. Using a suite of 55 PINOCCHIO simulations, with 2563 particles each, we study the MAHs of 12,924 cold dark matter haloes in a Λ CDM concordance cosmology. This is by far the largest set of haloes used for any such analysis. For each MAH we derive four different formation redshifts, which characterize different epochs during the assembly history of a dark matter halo. We show that haloes less massive than the characteristic non-linear mass scale establish their potential wells much before they acquire most of their mass. The time when a halo reaches its maximum virial velocity roughly divides its mass assembly into two phases, a fast accretion phase which is dominated by major mergers, and a slow accretion phase dominated by minor mergers. Each halo experiences about 3 ± 2 major mergers since its main progenitor had a mass equal to one percent of the final halo mass. This major merger statistic is found to be virtually independent of halo mass. However, the average redshift at which these major mergers occur is strongly mass dependent, with more massive haloes experiencing their major mergers later.

Key words: cosmology: theory — galaxies: formation — galaxies: haloes — dark matter.

1 INTRODUCTION

The cold dark matter (CDM) paradigm has become the standard framework for the formation of large-scale structure and galaxies. Small fluctuations in the initial density field grow by means of gravitational instability until they collapse to form virialized dark matter haloes. This growth process is hierarchical in the sense that small clumps virialize first, and aggregate successively into larger and larger objects. Galaxies form from the gas that is shock heated by the gravitational collapse and then subsequently cools (White & Rees 1978; but see also Birnboim & Dekel 2003 and Keres et al. 2004). Therefore, a proper understanding of galaxy formation relies on an accurate description of the structure and assembly of these dark matter haloes. This problem is tackled by a combination of both N-body simulations and analytical models. Although N-body simulations have the

advantage that they follow the formation of dark matter haloes into the non-linear regime, they are expensive, both in terms of labor (analyzing the simulations) and CPU time. Therefore, accurate analytical models are always useful. The most developed of these is the Press-Schechter (PS) formalism, which allows one to compute the (unconditional) halo mass function (Press & Schechter 1974). Bond et al. (1991), Bower (1991) and Lacey & Cole (1993) extended the PS formalism, using the excursion set approach, to compute conditional mass functions. These allow the construction of merger histories, the computation of halo formation times, and detailed studies of spatial clustering and large scale bias (e.g., Kauffmann & White 1993; Mo & White 1996, 2002; Mo, Jing & White 1996, 1997; Catelan et al. 1998; Sheth 1998; Nusser & Sheth 1999; Somerville & Kolatt 1999; Cohn, Bagla & White 2001).

Numerous studies in the past have tested the predictions of extended Press-Schechter (EPS) theory against numerical simulations. Although the unconditional mass func-

* E-mail: liyun@nova.astro.umass.edu

tion was found to be in reasonable agreement, it systematically over (under) predicts the number of low (high) mass haloes (e.g., Jain & Bertschinger 1994; Tormen 1998; Gross et al. 1998; Governato et al. 1999; Jenkins et al. 2001). Similar discrepancies have been found regarding the conditional mass function (Sheth & Lemson 1999; Somerville et al. 2000), which results in systematic offsets of the halo formation times predicted by EPS (e.g., van den Bosch 2002a). Finally, Bond et al. (1991) have shown that the PS approach achieves a very poor agreement on an object-by-object basis when compared with simulations (for a review, see Monaco 1998).

It is generally understood that these discrepancies stem from the assumption of spherical collapse. Numerous studies have investigated schemes to improve the EPS formalism by using ellipsoidal, rather than spherical collapse conditions, thereby taking proper account of the aspherical nature of collapse in a CDM density field (e.g., Sheth, Mo & Tormen 2001, hereafter SMT01; Sheth & Tormen 2002; Chiueh & Lee 2001; Lin, Chiueh & Lee 2002). Although this results in unconditional mass functions that are in much better agreement with numerical simulations (e.g., SMT01; Jenkins et al. 2001), they have been unable thus far to yield conditional mass functions of sufficient accuracy that reliable merger trees can be constructed.

Despite its systematic errors and uncertainties, the PS formalism has remained the standard analytical approach in galaxy formation modeling. In particular, the extended Press-Schechter theory is used extensively to compute merger histories and mass assembly histories (hereafter MAHs) which serve as the back-bone for models of galaxy formation (e.g., Kauffmann, White & Guiderdoni 1993; Somerville & Primack 1999; Cole et al. 2000; van den Bosch 2001; Firmani & Avila-Reese 2000). This may have profound implications for the accuracy of these models. For instance, the mass assembly histories of dark matter haloes are expected to impact on the star formation histories of the galaxies that form inside these haloes. In addition, the merger and mass assembly history of individual haloes may also be tightly related to their internal structure. As shown by Wechsler et al. (2002; hereafter W02) and Zhao et al. (2003a;b), the MAH is directly related to the concentration of the resulting dark matter halo (see also Navarro, Frenk & White 1997; Bullock et al. 2001; Eke, Navarro & Steinmetz 2001). Errors in the mass assembly histories of dark matter haloes may therefore result in erroneous predictions regarding the star formation history and the rotation curve shapes and/or the zero-point of the Tully-Fisher relation (e.g., Alam, Bullock & Weinberg 2002; Zentner & Bullock 2002; Mo & Mao (2000); van den Bosch, Mo & Yang 2003). Clearly, a detailed understanding of galaxy formation requires a description of the growth history of dark matter haloes that is more accurate than EPS. Although N -body simulations are probably the most reliable means of obtaining accurate assembly histories of dark matter haloes, they are computationally too expensive for some purposes.

As an alternative to the EPS formalism and N -body simulations, perturbative techniques have been developed that describe the growth of dark matter haloes in a given numerical realization of a linear density field. These include, amongst others, the truncated Zel'dovich (1970) approximation (Borgani, Coles & Moscardini 1994), the peak-patch al-

gorithm (Bond & Myers 1996a,b) and the merging cell model (Rodriguez & Thomas 1996; Lanzoni, Mamon & Guiderdoni 2000). Recently, Monaco, Theuns & Taffoni (2002b) developed a numerical code that uses local ellipsoidal collapse approximations (Bond & Myers 1996a; Monaco 1995) within Lagrangian Perturbation Theory (LPT, Buchert & Ehlers 1993; Catelan 1995). This code, called PINOCCHIO (PIN-pointing Orbit-Crossing Collapsed Hierarchy Objects), has been shown to yield accurate mass functions, both conditional and unconditional (Monaco et al. 2002a,b; Taffoni, Monaco & Theuns 2002), and is therefore ideally suited to study halo assembly histories, without having to rely on computationally expensive N -body simulations.

This paper is organized as follows. In Section 2 we give a detailed overview of (extended) Press-Schechter theory, including a discussion of its short-comings and its modifications under ellipsoidal collapse conditions, and describe the Lagrangian perturbation code PINOCCHIO. In Section 3 we compare the MAHs obtained from PINOCCHIO, the EPS formalism, and N -body simulations. We show that PINOCCHIO yields MAHs that are in excellent agreement with numerical simulations, and do not suffer from the shortcomings of the EPS formalism. In the second part of this paper we then analyze a large, statistical sample of MAHs obtained with PINOCCHIO for haloes spanning a wide range in masses. In Section 4 we use these MAHs to study, in a statistical sense, various characteristic epochs and events in the mass assembly history of a typical CDM halo. We analyze the statistics of major merger events in Section 5. Finally, Section 6 summarizes our results.

2 THEORETICAL BACKGROUND

2.1 Extended Press-Schechter theory

In the standard model for structure formation the initial density contrast $\delta(\mathbf{x}) = \rho(\mathbf{x})/\bar{\rho} - 1$ is considered to be a Gaussian random field, which is therefore completely specified by the power spectrum $P(k)$. As long as $\delta \ll 1$ the growth of the perturbations is linear and $\delta(\mathbf{x}, t_2) = \delta(\mathbf{x}, t_1)D(t_2)/D(t_1)$, where $D(t)$ is the linear growth factor linearly extrapolated to the present time. Once $\delta(\mathbf{x})$ exceeds a critical threshold δ_{crit}^0 the perturbation starts to collapse to form a virialized object (halo). In the case of spherical collapse $\delta_{\text{crit}}^0 \simeq 1.68$. In what follows we define δ_0 as the initial density contrast field linearly extrapolated to the present time. In terms of δ_0 , regions that have collapsed to form virialized objects at redshift z are then associated with those regions for which $\delta_0 > \delta_c(z) \equiv \delta_{\text{crit}}^0/D(z)$.

In order to assign masses to these collapsed regions, the PS formalism considers the density contrast δ_0 smoothed with a spatial window function (filter) $W(r; R_f)$. Here R_f is a characteristic size of the filter, which is used to compute a halo mass $M = \gamma_f \bar{\rho} R_f^3/3$, with $\bar{\rho}$ the mean mass density of the Universe and γ_f a geometrical factor that depends on the particular choice of filter. The *ansatz* of the PS formalism is that the fraction of mass that at redshift z is contained in haloes with masses greater than M is equal to two times the probability that the density contrast smoothed with $W(r; R_f)$ exceeds $\delta_c(z)$. This results in the well known

PS mass function for the comoving number density of haloes:

$$\frac{dn}{d \ln M}(M, z) dM = \sqrt{\frac{2}{\pi}} \frac{\delta_c(z)}{\sigma_2(M)} \left| \frac{d\sigma}{dM} \right| \exp \left[-\frac{\delta_c^2(z)}{2\sigma_2^2(M)} \right] dM \quad (1)$$

(Press & Schechter 1974). Here $\sigma_2(M)$ is the mass variance of the smoothed density field given by

$$\sigma_2(M) = \frac{1}{2\pi^2} \int_0^\infty P(k) \widehat{W}_2(k; R_f) k^2 dk. \quad (2)$$

with $\widehat{W}(k; R_f)$ the Fourier transform of $W(r; R_f)$.

The *extended* Press-Schechter (EPS) model developed by Bond et al. (1991), is based on the excursion set formalism. For each point one constructs ‘trajectories’ $\delta(M)$ of the linear density contrast at that position as function of the smoothing mass M . In what follows we adopt the notation of Lacey & Cole (1993) and use the variables $S = \sigma_2(M)$ and $\omega = \delta_c(z)$ to label mass and redshift, respectively. In the limit $R_f \rightarrow \infty$ one has that $S = \delta(S) = 0$, which can be considered the starting point of the trajectories. Increasing S corresponds to decreasing the filter mass M , and $\delta(S)$ starts to wander away from zero, executing a random walk (if the filter is a sharp k -space filter). The fraction of matter in collapsed objects in the mass interval $M, M+dM$ at redshift z is now associated with the fraction of trajectories that have their *first upcrossing* through the barrier $\omega = \delta_c(z)$ in the interval $S, S+dS$, which is given by

$$P(S, \omega) dS = \frac{1}{\sqrt{2\pi}} \frac{\omega}{S^{3/2}} \exp \left[-\frac{\omega^2}{2S} \right] dS \quad (3)$$

(Bond et al. 1991; Bower 1991; Lacey & Cole 1993). After conversion to number counting, this probability function yields the PS mass function of equation (1). Note that this approach does not suffer from the arbitrary factor two in the original Press & Schechter approach.

Since for random walks the upcrossing probabilities are independent of the path taken (i.e., the upcrossing is a Markov process), the probability for a change ΔS in a time step $\Delta\omega$ is simply given by equation (3) with S and ω replaced with ΔS and $\Delta\omega$, respectively. This allows one to immediately write down the *conditional* probability that a particle in a halo of mass M_2 at z_2 was embedded in a halo of mass M_1 at z_1 (with $z_1 > z_2$) as

$$P(S_1, \omega_1 | S_2, \omega_2) dS_1 = \frac{1}{\sqrt{2\pi}} \frac{(\omega_1 - \omega_2)}{(S_1 - S_2)^{3/2}} \exp \left[-\frac{(\omega_1 - \omega_2)^2}{2(S_1 - S_2)} \right] dS_1 \quad (4)$$

Converting from mass weighting to number weighting, one obtains the average number of progenitors at z_1 in the mass interval M_1, M_1+dM_1 which by redshift z_2 have merged to form a halo of mass M_2 :

$$\frac{dN}{dM_1}(M_1, z_1 | M_2, z_2) dM_1 = \frac{M_2}{M_1} P(S_1, \omega_1 | S_2, \omega_2) \left| \frac{dS}{dM} \right| dM_1. \quad (5)$$

This conditional mass function can be combined with Monte-Carlo techniques to construct merger histories (also called merger trees) of dark matter haloes.

2.2 Ellipsoidal collapse

In an attempt to improve the inconsistencies between EPS and numerical simulations (see Section 1), various authors have modified the EPS formalism by considering ellipsoidal rather than spherical collapse. For ellipsoidal density perturbations, the conditions for collapse not only depend on the self-gravity of the perturbation, but also on the tidal coupling with the external mass distribution; external shear can actually rip overdensities apart and thus prevent them from collapsing. Since smaller mass perturbations typically experience a stronger shear field, they tend to be more ellipsoidal. Therefore, it is to be expected that the assumptions of spherical collapse in the standard EPS formalism are more accurate for more massive haloes, whereas modifications associated with ellipsoidal collapse will be more dramatic for smaller mass haloes. The way in which ellipsoidal collapse modifies the halo formation times with respect to the EPS predictions depends on the definition of collapse. Ellipsoidal perturbations collapse independently along the three different directions defined by the eigen vectors of the deformation tensor (defined as the second derivative of the linear gravitational potential). It is customary to associate the first axis collapse with the formation of a 2-dimensional pancake-like structure, the second axis collapse with the formation of a 1-dimensional filament, and the third axis collapse with the formation of a dark matter halo. Most authors indeed have associated halo formation with the collapse of the third axis (e.g., Bond & Myers 1996a; Audit, Teyssier & Alimi 1997; Lee & Shandarin 1998; SMT01), though some have considered the first axis collapse instead (e.g., Bertschinger & Jain 1994; Monaco 1995). For first-axis collapse one predicts that haloes form earlier than in the spherical case, whereas the opposite applies when considering third-axis collapse. Clearly, the implications of considering ellipsoidal rather than spherical collapse depend sensitively on the collapse definition.

In order to incorporate ellipsoidal collapse in a PS-like formalism, one needs to obtain an estimate of the critical overdensity for collapse δ_{ec} . Various studies have attempted such schemes. For instance, SMT01 used the ellipsoidal collapse model to obtain

$$\delta_{ec}(M, z) = \delta_c(z) \left(1 + 0.47 \left[\frac{\sigma_2(M)}{\delta_{c2}(z)} \right]^{0.615} \right). \quad (6)$$

Here $\delta_c(z)$ is the standard value for the spherical collapse model. Solving for the upcrossing statistics with this particular barrier shape results in halo mass functions that are in excellent agreement with those found in simulations (Sheth & Tormen 1999; Jenkins et al. 2001). Unfortunately, no analytical expression for the conditional mass function is known for a barrier of the form of equation (6), and one has to resort to either approximate fitting functions (Sheth & Tormen 2002), or one has to use time-consuming Monte-Carlo simulations to determine the upcrossing statistics (Chiueh & Lee 2001; Lin et al. 2002). Although the resulting conditional mass functions $\frac{dN}{dM_1}(M_1, z_1 | M_2, z_2) dM_1$ have been found to be in good agreement with numerical simulations if a relatively large look-back time is considered (i.e., if $\Delta z = z_2 - z_1 \gtrsim 0.5$), there is still a large disagreement for small Δz . This is probably due to the neglect of correlations between scales in the excursion set approach (Peacock

& Heavens 1990; Sheth & Tormen 2002). This is unfortunate as it does not allow these methods to be used for the construction of merger histories or MAHs. Lin et al. (2002) tried to circumvent this problem by introducing a small mass gap between parent halo and progenitor halo, i.e., each time step they require that $S_1 - S_2 \geq f \delta_c 2(z_2)$. Upon testing their conditional mass function with this mass gap against numerical simulations they find good agreement for $f = 0.06$, and claim that with this modification the excursion set approach *can* be used to construct merger histories under ellipsoidal collapse conditions. However, they only tested their conditional mass functions for $\Delta z \geq 0.2$, whereas accurate merger histories require significantly smaller time steps. For instance, van den Bosch (2002a) has argued for timesteps not larger than $\Delta\omega = \omega_1 - \omega_2 \simeq 0.1$, which, for an Einstein-de Sitter (EdS) cosmology, corresponds to $\Delta z \simeq 0.06$ (see also discussion in Somerville & Kolatt 1999). Furthermore, with the mass gap suggested by Lin et al. (2002), each time step there is a minimum amount of mass accreted by the halo, which follows from $S_1 - S_2 = f \delta_c 2(z_2)$. This introduces a distinct maximum to the halo half-mass formation time, the value of which depends sensitively on the actual time-steps taken. To test this, we constructed MAHs of CDM haloes using the method of van den Bosch (2002a) but adopting the conditional probability function of Lin et al. (2002). This resulted in MAHs that are in very poor agreement with numerical simulations. In particular, the results were found to depend strongly on the value of $\Delta\omega$ adopted.

In summary, although introducing ellipsoidal collapse conditions in the excursion set formalism has allowed the construction of accurate unconditional mass functions, there still is no reliable method based on the EPS formalism that allows the construction of accurate merger histories and/or MAHs.

2.3 PINOCCHIO

Although the problem of obtaining accurate merging histories under ellipsoidal collapse conditions can be circumvented by using N-body simulations, the time-expense of these simulations is a major hurdle. An attractive alternative is provided by the LPT code PINOCCHIO developed recently by Monaco et al. (2002b). Below we give a short overview of PINOCCHIO, and we refer the interested reader to Monaco et al. (2002a,b) and Taffoni et al. (2002) for a more elaborate description.

PINOCCHIO uses Lagrangian perturbation theory to describe the dynamics of gravitational collapse. In LPT the comoving (Eulerian) coordinate \mathbf{x} and the initial Lagrangian coordinate \mathbf{q} of each particle are connected via

$$\mathbf{x}(\mathbf{q}, t) = \mathbf{q} + \mathbf{S}(\mathbf{q}, t), \quad (7)$$

with \mathbf{S} the displacement field. The first-order term of $\mathbf{S}(\mathbf{q}, t)$ is the well-known Zel'dovich approximation (Zel'dovich 1970):

$$\mathbf{S}(\mathbf{q}, t) = -D(t) \frac{\partial \psi}{\partial \mathbf{q}} \quad (8)$$

with $\psi(\mathbf{q})$ the rescaled linear gravitational potential, which is related to the density contrast $\delta_0(\mathbf{q})$ extrapolated to the present time by the Poisson equation

$$\nabla^2 \psi(\mathbf{q}) = \delta_0(\mathbf{q}), \quad (9)$$

Since the Lagrangian density field is basically $\rho_L(\mathbf{q}) = \bar{\rho}$, the (Eulerian) density contrast is given by

$$1 + \delta(\mathbf{x}, t) = \frac{1}{\det(J)} \quad (10)$$

with $J = \partial \mathbf{x} / \partial \mathbf{q}$ the Jacobian of the transformation given in (7). Note that the density formally goes to infinity when the Jacobian determinant vanishes, which corresponds to the point in time when the mapping $\mathbf{q} \rightarrow \mathbf{x}$ becomes multi-valued, i.e. when orbits first cross leading to the formation of a caustic. Since the (gravitationally induced) flow is irrotational the matrix J is symmetric and can thus be diagonalized:

$$1 + \delta(\mathbf{x}, t) = \frac{1}{\prod_{i=1}^3 [1 - D(t) \lambda_i(\mathbf{q})]} \quad (11)$$

with $-\lambda_i$ the eigenvalues of the deformation tensor $\partial^2 \psi / \partial q_i \partial q_j$.

PINOCCHIO starts by constructing a random realization of a Gaussian density field $\rho(\mathbf{q})$ (linearly extrapolated to $z = 0$) and the corresponding peculiar potential $\phi(\mathbf{q})$ on a cubic grid. The density fluctuation field is specified completely by the power spectrum $P(k)$, which is normalized by specifying the value of σ_8 , defined as the rms linear overdensity at $z = 0$ in spheres of radius $8h^{-1}$ Mpc. The density and peculiar potential fields are subsequently convolved with a series of Gaussians with different values for their FWHM R . For the 2563 simulations used in this paper, 26 different linearly sampled values of R are used. For a given value of R the density of a mass element (i.e., ‘particle’) will become infinite as soon as at least one of the ellipsoid’s axes reaches zero size (i.e., when $D(t) = 1/\lambda_i$). At this point orbit crossing (OC) occurs and the mass element enters a high-density multi-stream region. This is the moment of first-axis collapse. Since the Jacobian determinant becomes multivalued at this stage, one can not make any further predictions of the mass element’s fate beyond this point in time. Consequently, it is not possible in PINOCCHIO to associate halo collapse with that of the third axis.

For each Lagrangian point \mathbf{q} (hereafter ‘particle’) and for each smoothing radius R this OC (i.e., collapse) time is computed, and the highest collapse redshift z_c , the corresponding smoothing scale R_c , and the Zel’dovich estimate of the peculiar velocity \mathbf{v}_c are recorded. PINOCCHIO differs from the standard PS-like method when it comes to assigning masses to collapsed objects. Rather than associating a halo mass with the collapsed mass element based directly on the smoothing scale R_c at collapse, PINOCCHIO uses a fragmentation algorithm to link neighboring mass elements into a common dark matter halo. In fact, the collapsed mass element may be assigned to a filament or sheet rather than a halo.

After sorting particles according to decreasing collapse redshift z_c the following rules for accretion and merging are adopted: Whenever a particle collapses and none of its Lagrangian neighbors (the six nearest particles) have yet collapsed, the particle is considered a seed for a new halo. Otherwise, the particle is accreted by the nearest Lagrangian neighbor that already has collapsed if the Eulerian distance d , computed using the Zel’dovich velocities \mathbf{v} at the time of collapse, obeys $d \leq f_a R_M$, where $R_M = M^{1/3}$ is the ra-

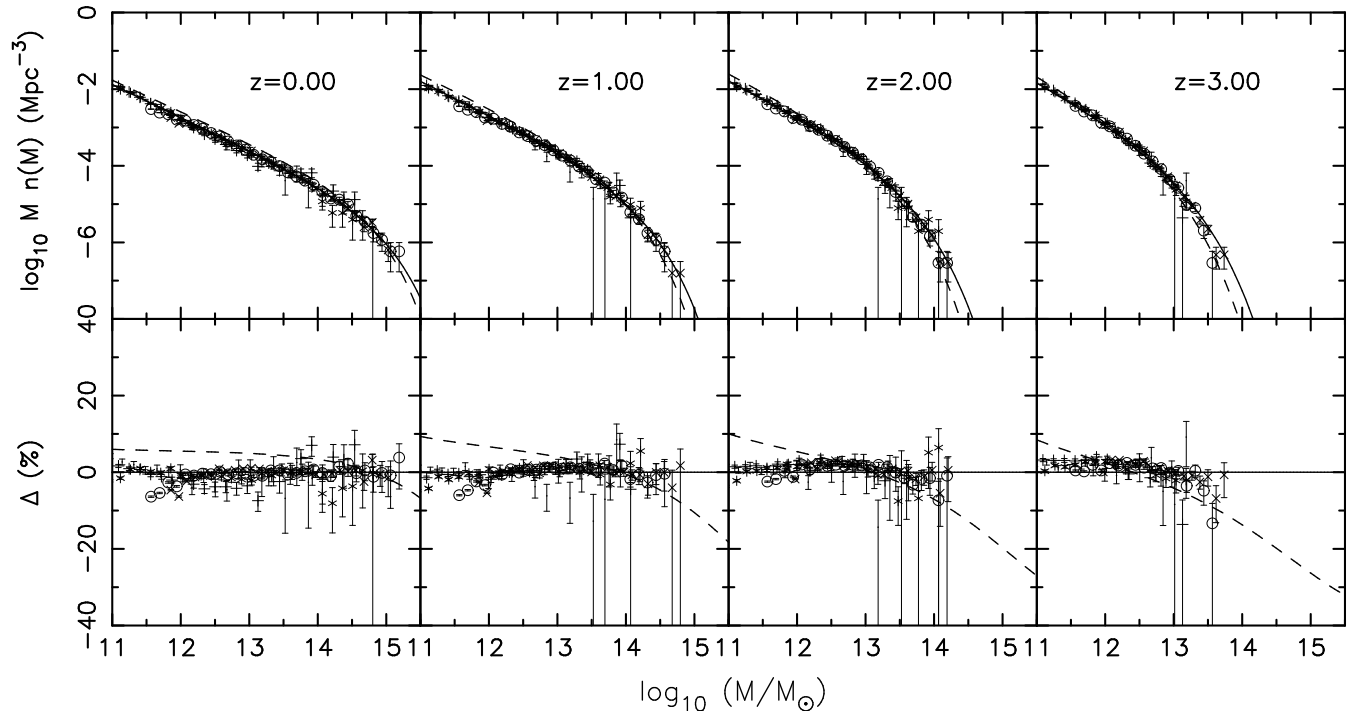


Figure 1. Panels in the upper row show the (unconditional) halo mass functions at 4 different redshifts, as indicated. Different symbols (each with Poissonian error bars) correspond to 5 different PINOCCHIO simulations randomly selected from P0, each with a different mass resolution. Dashed and solid lines correspond to the PS and SMT01 mass functions, respectively, and are shown for comparison. Panels in the lower row show the percentual difference between the PS and SMT01 mass functions (dashed lines) and that between the PINOCCHIO and the SMT01 mass functions (symbols with errorbars). Clearly, the PS mass function overestimates (underestimates) the number of small (high) mass haloes, while PINOCCHIO yields mass functions that are in excellent agreement with SMT01 (and thus with N-body simulations). Note that the SMT01 halo mass function best fits the mass function of simulated halos that is identified with an FOF linking length of 0.2 times the mean particle separation. The mean density of a halo so selected is similar to that within a virialized halo based on the spherical collapse model. PINOCCHIO haloes and PS haloes are all defined so that the mean density within a halo is similar to that based on the spherical collapse model.

dius of a halo of M particles. If more than one Lagrangian neighbor has already collapsed, it is simultaneously checked whether these haloes merge. This occurs whenever, again at the time of collapse, the mutual Eulerian distance between these haloes is $d \leq f_M R_M$, where R_M refers to the larger halo. Note that with this description, up to six haloes may merge at a given time. The collapsing particles that according to these criteria do not accrete onto a halo at their collapse time are assigned to a filament. In order to mimic the accretion of filaments onto haloes, filament particles can be accreted by a dark matter halo at a later stage when they neighbor (in Lagrangian space) an accreting particle. Finally, in high density regions it can happen that pairs of haloes that are able to merge are not touched by newly collapsing particles for a long time. Therefore, at certain time intervals pairs of touching haloes are merged if they obey the above merging condition.

The accretion and merging algorithm described above has five free parameters. In addition to the parameters f_a and f_M three additional free parameters have been introduced by Monaco et al. (2002b). We refer the reader to this paper for details. This relatively large amount of freedom may seem a weakness of PINOCCHIO. However, it is important to realize that even N-body codes require some free parameters, such as the linking-length in the Friends-Of-Friends (FOF) algorithm used to identify dark matter

haloes. Furthermore, we do not consider these parameters as free in what follows. Rather, we adopt the values advocated by Monaco et al. (2002a,b), which they obtained by tuning PINOCCHIO to reproduce the conditional and unconditional mass function of N-body simulations.

3 SIMULATIONS

In this paper we use PINOCCHIO simulations to study the mass assembly histories (MAHs) of dark matter haloes. We follow previous studies (Lacey & Cole 1993; Eisenstein & Loeb 1996; Nusser & Sheth 1999; van den Bosch 2002a) and define the MAH, $M(z)$, of a halo as the main trunk of its merger tree: at each redshift, the mass $M(z)$ is associated with the mass of the most massive progenitor at this redshift, and we follow this progenitor, and this progenitor only, further back in time. In this way, this ‘main progenitor halo’ never accretes other haloes that are more massive than itself. Note that although at each branching point we follow the most massive branch, this does not necessarily imply that the main progenitor is also the most massive of all progenitors at any given redshift.

Below we describe the PINOCCHIO simulations, the N-body simulations, and the EPS method used to construct MAHs.

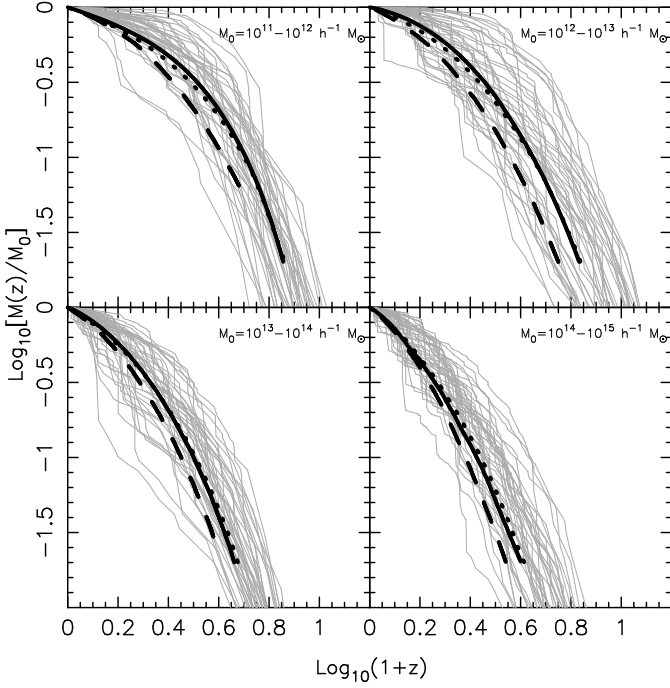


Figure 2. The mass assembly histories of dark matter haloes with present-day masses in the four mass bins as indicated in the panels. The upper two panels are based on the $100h^{-1}\text{Mpc}$ -box simulations, P1 and S1, while the lower two panels use data from the $300h^{-1}\text{Mpc}$ -box simulations, P2 and S2. The thin lines are 40 MAHs randomly selected from the PINOCCHIO simulations. The thick solid line in each panel shows the average of all the MAHs obtained in the PINOCCHIO simulations in the corresponding mass bin. The thick dotted line shows the average MAH extracted from the simulations. The thick dashed line shows the average MAH obtained from 3000 EPS realizations (properly sampled from halo mass function).

Table 1. Ensemble of PINOCCHIO simulations (P0)

Box size (h^{-1} Mpc)	N_{run}	M_p ($h^{-1} M_{\odot}$)	N_{MAH}
20	12	4.0×10^7	2,690
40	8	3.2×10^8	1,863
60	8	1.1×10^9	796
80	6	2.5×10^9	1,438
100	6	5.0×10^9	2,799
140	4	1.4×10^{10}	410
160	2	2.0×10^{10}	299
200	9	4.0×10^{10}	2,629

A listing of the PINOCCHIO simulations used in this paper. All simulations use 2563 particles and adopt the standard ΛCDM concordance cosmology. In order to get good statistics, we choose a combination of box sizes so that we can select thousands of well-resolved (with more than 2000 particles) haloes in each mass bin we adopt in the paper. This ensemble of PINOCCHIO simulations is referred to as ‘P0’ in the text. The first column of Table 1 lists the box size of the simulation in h^{-1} Mpc. The second column lists the number of independent realizations run. The particle mass M_p (in $h^{-1} M_{\odot}$) is listed in the third column, while the fourth column lists the total number of haloes (summed over all N_{run} realizations) with more than 2000 particles and for which a MAH has been obtained.

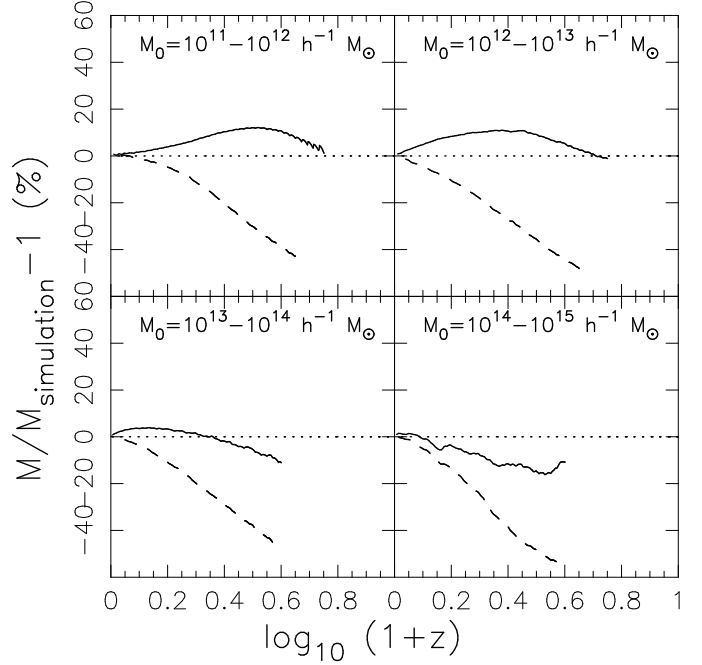


Figure 3. The dashed curve in each panel shows the difference between the average MAHs predicted by the EPS model and by the N-body simulation, while the solid curve shows the difference between PINOCCHIO prediction and N-body simulation. The upper two panels use data from P1 and S1, while the lower two panels use data from P2 and S2. Data are not shown for $z \gtrsim 3$ because the MAHs are not well represented at such high redshifts in the simulations.

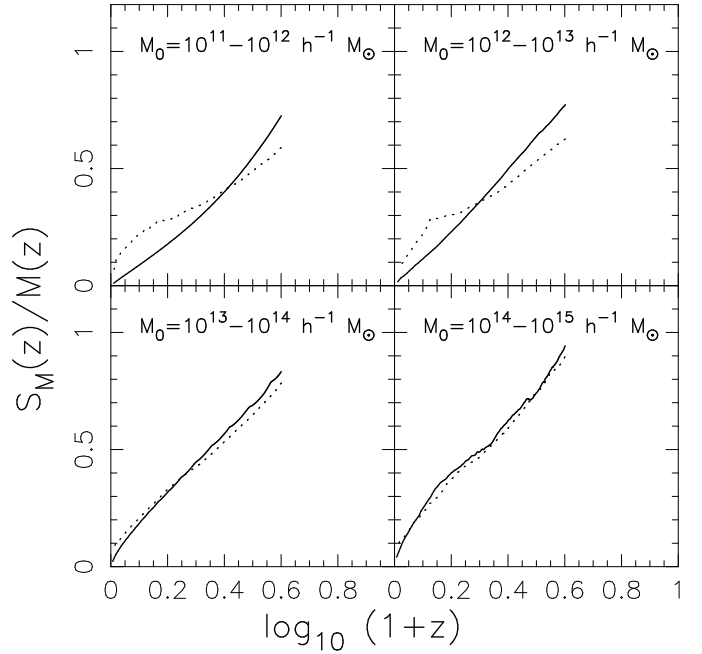


Figure 4. The standard deviation of the MAHs, $S_M(z)$, normalized by the average MAH, $M(z)$, in four mass bins. Solid lines are results from PINOCCHIO, while dotted lines are results from N-body simulations. As in Fig. 2 and Fig. 3, the upper two panels use data from P1 and S1, while the lower two panels use data from P2 and S2.

3.1 PINOCCHIO simulations

Because the progenitors of a present-day halo become smaller at higher redshift, we can only follow the MAHs to a sufficiently high redshift if the halo at $z = 0$ contains a large enough number of particles. When constructing MAHs with PINOCCHIO, we only use haloes that contain more than 2000 particles at the present time, and we trace each MAH to the redshift at which its main progenitor contains less than 10 particles. In order to cover a large range of halo masses, we have carried out 55 PINOCCHIO simulations with 2563 particles each and spanning a wide range of box sizes and particle masses (see Table 1, we call this suite of PINOCCHIO simulations P0 hereafter). The choice of box sizes ensures that there are several thousand well-resolved haloes in each of the mass bins considered. Each of these simulations takes only about 6 hours of CPU time on a common PC (including the actual analysis), clearly demonstrating its advantage over regular N-body simulations. This suite of PINOCCHIO simulations has adopted the Λ CDM concordance cosmology with $\Omega_m = 0.3$, $\Omega_\Lambda = 0.7$, $h = 0.7$ and $\sigma_8 = 0.9$.

With simulation box sizes ranging from $20 h^{-1}\text{Mpc}$ to $200 h^{-1}\text{Mpc}$, and particle masses ranging from $4 \times 10^7 h^{-1} M_\odot$ to $4 \times 10^{10} h^{-1} M_\odot$, we are able to study the MAHs of present-day haloes with masses $> 8 \times 10^{10} h^{-1} M_\odot$. The construction of the MAHs is straightforward: PINOCCHIO outputs a halo mass every time a merger occurs, i.e., when a halo with more than 10 particles merges into the main branch. If we require an estimate of the halo mass at any intermediate redshift, z , we use linear interpolation in $\log(1+z)$ between the two adjacent output redshifts.

3.2 N-body simulations

For comparison we also used MAHs extracted from two sets of N-body simulations (referred to as S1 and S2). These N-body simulations follow the evolution of 5123 particles in a periodic box of $100 h^{-1}\text{Mpc}$ (S1) and $300 h^{-1}\text{Mpc}$ (S2) on a side, assuming slightly different cosmologies (see Table 2 for details). The snapshot outputs of each simulation are evenly placed at 60 redshifts between $z = 0$ and $z = 15$ in $\ln(1+z)$ space.

In each simulation and at each output, haloes are identified using the standard FOF algorithm with a linking length of $b = 0.2$. Haloes obtained with this linking length have a mean overdensity of ~ 180 . A halo at redshift z_1 is identified as a progenitor of a halo at $z_2 < z_1$ if more than half of its mass is included in the halo at z_2 . The resulting lists of progenitor haloes are used to construct the MAHs. In our analysis, we only use haloes more massive than $10^{11} h^{-1} M_\odot$ at the present time in S1 and halos more massive than $10^{13} h^{-1} M_\odot$ in S2. Thus, in each simulation only halos with more than ~ 600 particles at $z = 0$ are used, which allows us to trace the MAHs to sufficiently high redshift with sufficiently high resolution. For comparison, we also generate two sets of PINOCCHIO simulations, P1 and P2, using exactly the same numbers of particles and cosmologies as in S1 and S2, respectively (see Table 2).

3.3 Monte-Carlo simulations

We also generate MAHs using Monte-Carlo simulations based on the standard EPS formalism. We adopt the N-branch tree method with accretion suggested by Somerville & Kolatt (1999, hereafter SK99). This method yields more reliable MAHs than for example the binary-tree method of Lacey & Cole (1993). In particular, it ensures exact mass conservation, and yields conditional mass functions that are in good agreement with direct predictions from EPS theory (i.e., the method is self-consistent).

To construct a merger tree for a parent halo of mass M the SK99 method works as follows. First a value for ΔS is drawn from the mass-weighted probability function

$$P(\Delta S, \Delta\omega) d\Delta S = \frac{1}{\sqrt{2\pi}} \frac{\Delta\omega}{\Delta S^{3/2}} \exp\left[-\frac{(\Delta\omega)^2}{2\Delta S}\right] d\Delta S \quad (12)$$

(cf. equation [4]). Here $\Delta\omega$ is a measure for the time step used in the merger tree, and is a free parameter (see below). The progenitor mass, M_p , corresponding to ΔS follows from $\sigma^2(M_p) = \sigma^2(M) + \Delta S$. With each new progenitor it is checked whether the sum of the progenitor masses drawn thus far exceeds the mass of the parent, M . If this is the case the progenitor is rejected and a new progenitor mass is drawn. Any progenitor with $M_p < M_{\min}$ is added to the mass component M_{acc} that is considered to be accreted onto the parent in a smooth fashion (i.e., the formation history of these small mass progenitors is not followed further back in time). Here M_{\min} is a free parameter that has to be chosen sufficiently small. This procedure is repeated until the total mass left, $M_{\text{left}} = M - M_{\text{acc}} - \sum M_p$, is less than M_{\min} . This remaining mass is assigned to M_{acc} and one moves on to the next time step. For the construction of MAHs, however, it is not necessary to construct an entire set of progenitors. Rather, at each time step, one can stop once the most massive progenitor drawn thus far is more massive than M_{left} . This has the additional advantage that one does not have to define a minimum progenitor mass M_{\min} (see van den Bosch 2002a for details).

In principle, since the upcrossing of trajectories through a boundary is a Markov process, the statistics of progenitor masses should be independent of the time steps taken. However, the SK99 algorithm is based on the *single* halo probability (equation [12]), which does not contain any information about the *set* of progenitors that make up the mass of M . In fact, mass conservation is enforced ‘by hand’, by rejecting progenitor masses that overflow the mass budget. As shown in van den Bosch (2002a), this results in a time step dependency, but only for relatively large time steps. For sufficiently small values of $\Delta\omega$ the algorithm outlined above yields accurate and robust results (see also SK99). Throughout this paper we adopt a timestep of $\Delta z = 0.05$. Our tests with different values of Δz from 0.01 to 0.05 have shown that this time step is small enough to achieve stable results, that is, when we decrease the time step to $\Delta z = 0.01$, the change in the average MAH is less than 1%.

3.4 Comparison

We now compare the MAHs obtained with all three methods discussed above. The upper panels of Fig. 1 plot the (unconditional) halo mass functions at four different redshifts,

Table 2. Reference PINOCCHIO and N-body simulations

Simulation Name	N_p	Box size (h^{-1} Mpc)	$M_p(h^{-1} M_\odot)$	Ω_m	Ω_Λ	h	σ_8
S1 (N-body)	5123	100	5.5×10^8	0.268	0.732	0.71	0.85
P1 (PINOCCHIO)	5123	100	5.5×10^8	0.268	0.732	0.71	0.85
S2 (N-body)	5123	300	1.3×10^{11}	0.236	0.764	0.73	0.74
P2 (PINOCCHIO)	5123	300	1.3×10^{11}	0.236	0.764	0.73	0.74

as indicated, obtained from 5 arbitrary PINOCCHIO runs with different box sizes in P0. Dashed lines correspond to the analytical halo mass functions obtained using the standard PS formalism (equation [1]), while the solid lines indicate the mass functions of SMT01 based on ellipsoidal collapse. The latter have been shown to accurately match the mass functions obtained from N-body simulations (e.g., Sheth & Tormen, 1999; SMT01). The symbols in the lower panels of Fig. 1 plot the differences between the PINOCCHIO and the SMT01 mass functions, while the dashed lines indicate the differences between the PS and the SMT01 mass functions. Clearly, the PINOCCHIO mass functions are in excellent agreement with those of SMT01, and thus also with those obtained from N-body simulations. In addition, Taffoni et al. (2002) have shown that PINOCCHIO also accurately matches the *conditional* mass functions obtained from numerical simulations. We now investigate whether the actual MAHs obtained from PINOCCHIO are also in good agreement with the numerical simulations.

Fig. 2 plots the average MAHs obtained from the PINOCCHIO, N-body and EPS simulations, for halos with the present masses in the following four mass ranges: $\log(M_0/h^{-1} M_\odot) = 11-12, 12-13, 13-14$ and $14-15$. For comparison, in each panel we also show 40 randomly selected MAHs from the PINOCCHIO simulations (P1 and P2). To ensure mass resolution, results for the low-mass bins (the two upper panels) are based on simulations with the small box size, i.e. S1 and P1. Results for the high-mass bins (the two lower panels) are based only on simulations with the large-box size (S2 and P2) in order to obtain a large number of massive halos. The thick solid curve in each panel corresponds to the average MAH obtained by averaging over all the halos, in the mass range indicated, found in one of the PINOCCHIO simulations (P1 and P2). The thick dashed lines correspond to the average MAHs obtained from 3000 EPS Monte-Carlo simulations (properly weighted by the halo mass function). The thick dotted lines show the average MAHs obtained from the two N-body simulations (S1 and S2). In Fig. 3, a detailed comparison between these results are presented. As can be seen in Fig. 3, the average MAHs obtained with PINOCCHIO are in good agreement with those obtained from the N-body simulations (with differences smaller than 10%). Note that there are uncertainties in the identification of dark haloes in N-body simulations using the FOF algorithm. Sometimes two physically separated haloes can be linked together and identified as one halo if they are bridged by dark matter particles, which can change the halo mass by 5% on average. The agreement between PINOCCHIO and simulation shown in Fig. 3 is probably as good as one can hope for. The EPS model, however, yields

MAHs that are systematically offset with respect to those obtained from the N-body simulations: the EPS formalism predicts that haloes assemble too late (see also van den Bosch 2002a; Lin, Jing & Lin 2003; W02). Fig. 4 shows the ratio between the standard deviation of the MAHs, $S_M(z)$, and the average MAH $M(z)$, as a function of redshift z . As one can see, the agreement between the PINOCCHIO and N-body simulations is also reasonably good.

In summary, the Lagrangian Perturbation code PINOCCHIO yields halo mass functions (both conditional and unconditional), and mass assembly histories that are all in good agreement with N-body simulations. In particular, it works much better than the standard PS formalism, and yet is much faster to run than numerical simulations. PINOCCHIO therefore provides a unique and fast platform for accurate investigations of the assembly histories of a large, statistical sample of CDM haloes.

4 HALO FORMATION TIMES

Having demonstrated that the PINOCCHIO MAHs are in good agreement with those obtained from N-body simulations, we now use the suite of 55 PINOCCHIO simulations, P0, listed in Table 1 to investigate the assembly histories of a large sample of haloes spanning a wide range in halo masses.

The assembly history of a halo can be parameterized by a formation time (or equivalently formation redshift), which characterizes when the halo assembles. However, since the assembly of a halo is a continuous process, different ‘formation times’ can be defined, each focusing on a different aspect of the MAH. Here we define and compare the following four formation redshifts:

- (i) z_{half} : This is the redshift at which the halo has assembled half of its final mass. This formation time has been widely used in the literature.
- (ii) z_{imm} : This is redshift at which the halo experiences its last major merger. Unless stated otherwise we define a major merger as one in which the mass ratio between the two progenitors is larger than 1/3. This definition is similar to z_{jump} defined in Cohn & White (2005). Major mergers may have played an important role in transforming galaxies and in regulating star formation in galaxies. Their frequency is therefore important to quantify.
- (iii) z_{vvir} : This is the redshift at which the virial velocity of a halo, V_{vir} , defined as the circular velocity at the virial radius, reaches its current value, V_0 , for the first time. Since V_{vir} is a measure for the depth of the potential well, z_{vvir}

characterizes the formation time of the halo’s gravitational potential.

(iv) z_{vmax} : This is the redshift at which the halo’s virial velocity reaches its maximum value over the entire MAH. As we show below, the value of V_{vir} is expected to increase (decrease) with time, if the time scale for mass accretion is shorter (longer) than the time scale of the Hubble expansion. Therefore, z_{vmax} indicates the time when the MAH transits from a fast accretion phase to a slow accretion phase.

In an N-body simulation one can infer the virial velocity of a halo from its internal structure. In the case of PINOCCHIO simulations, however, no information regarding the density distribution of haloes is available. However, we may use the fact that CDM haloes always have a particular (redshift and cosmology dependent) overdensity. This allows us to define the virial velocity at redshift z as

$$V_{\text{vir}}(z) = \sqrt{\frac{GM_{\text{vir}}}{R_{\text{vir}}}} = \left[\frac{\Delta_{\text{vir}}(z)}{2} \right]^{1/6} [M_{\text{vir}}(z) H(z)]^{1/3} \quad (13)$$

Here M_{vir} and R_{vir} are the virial mass and virial radius of the halo, respectively, and $H(z)$ is the Hubble parameter. The quantity $\Delta_{\text{vir}}(z)$ is the density contrast between the mean density of the halo and the critical density for closure, for which we use the fitting formula of Bryan & Norman (1998),

$$\Delta_{\text{vir}}(z) = 18\pi^2 + 82[\Omega_m(z) - 1] - 39[\Omega_m(z) - 1]^2 \quad (14)$$

As an illustration, Fig. 5 plots the MAH, $M(z)/M_0$ (upper panel), and the history of the virial velocity, $V_{\text{vir}}(z)/V_0$ (lower panel) for a randomly selected halo (with $M_0 = 1.02 \times 10^{13} h^{-1} M_\odot$). All major merger events are marked by a solid dot plus arrow. The last major merger occurs at $z_{\text{lmm}} = 1.60$. The other formation redshifts, $z_{\text{half}} = 1.59$, $z_{\text{vvir}} = 3.77$, and $z_{\text{vmax}} = 1.23$ are marked by an open circle, a cross, and an open triangle, respectively.

Fig. 6 plots the correlations between the various formation redshifts, for haloes with masses in the range $10^{11} - 10^{12} h^{-1} M_\odot$. The value of r_s in each panel shows the corresponding Spearman rank-order correlation coefficients. Clearly, there is significant correlation among all the formation redshifts, but the scatter is quite large. This demonstrates that these different formation times characterize different aspects of a given MAH. Unlike simulation which outputs snapshots at arbitrary times, PINOCCHIO only outputs when a merger occurs and the merger is treated as instantaneous. Consequently, some formation times can have exactly the same value in PINOCCHIO simulations. Note that the correlation shown in the lower left panel is quite similar to that obtained by Cohn & White (2005) for simulated clusters of galaxies. Note also that typically, $z_{\text{vvir}} > z_{\text{half}}$ and $z_{\text{vvir}} \gg z_{\text{lmm}}$. This shows that haloes in this mass range established their potential wells before they accreted a major fraction of their mass. The last major merger typically occurred well before z_{half} , which indicates that most of that mass has been accreted in a fairly smooth fashion (see also W02 and Zhao et al. 2003a).

Fig. 7 shows the distributions of the four formation redshifts defined above. Results are shown for four different mass bins, as indicated. For all four formation redshifts, the median is higher for haloes of lower masses. This reflects the hierarchical nature of the assembly of dark matter haloes:

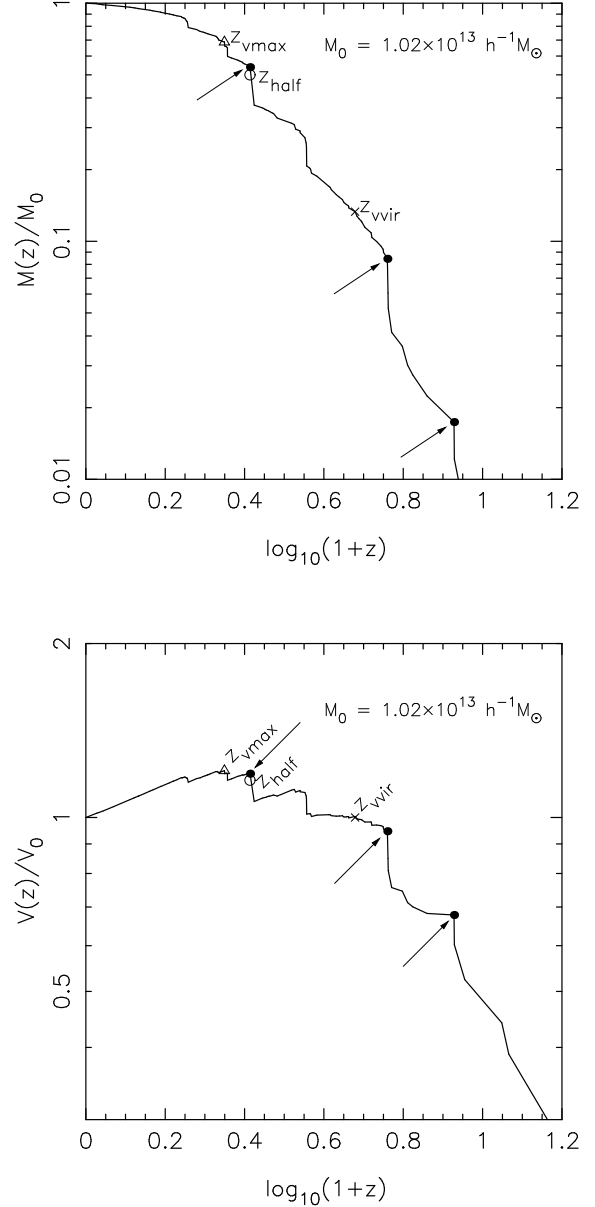


Figure 5. *Upper panel:* the MAH of a randomly chosen halo with a mass of $1.02 \times 10^{13} h^{-1} M_\odot$. Various characteristic events during the assembly of this halo are indicated: z_{vmax} (open triangle), z_{half} (open circle), and z_{vvir} (cross). The solid dots with an arrow indicate major mergers (those with a mass ratio larger than $1/3$). *Lower panel:* same as in upper panel, except that here the evolution of the halo virial velocity is shown.

less massive systems assemble (‘form’) earlier. Note that the distribution of formation times is also broader for lower mass haloes. For haloes with $M_0 \gtrsim M^* \simeq 10^{13} h^{-1} M_\odot$ ¹, all the distribution functions except that of z_{half} are peaked at, or very near to, $z = 0$. This shows that the majority of these haloes are still in their fast accretion phase, so that their

¹ Here M^* is the characteristic non-linear mass defined by $\sigma(M^*) = \delta_{\text{crit}0}$

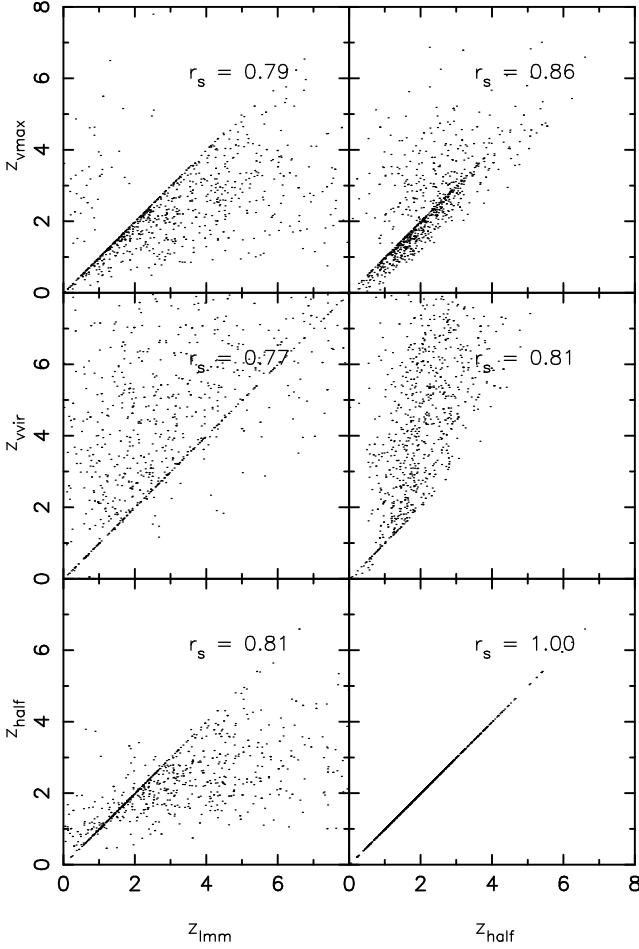


Figure 6. The correlations between various halo formation redshifts for haloes with present day masses in the range $10^{11} h^{-1} M_{\odot} \leq M \leq 10^{12} h^{-1} M_{\odot}$. The value of r_s in each panel shows the corresponding Spearman rank-order correlation coefficient. Due to the finite time resolution in the PINOCCHIO simulations, in some cases the values of two formation times can be the same.

potential wells are still deepening with time. On the other hand, haloes with $M_0 \ll M^*$ typically have $z_{\text{vvir}} > z_{\text{half}}$ and $z_{\text{vvir}} > z_{\text{lmm}}$ (cf. Fig. 6), indicating that their potential wells have already been established, despite the fact that they continue to accrete appreciable amounts of mass.

Fig. 8 shows the distributions of the ratio $M(z_{\text{form}})/M_0$, with z_{form} one of our four formation redshifts. By definition, the distribution of $M(z_{\text{half}})/M_0$ is a δ -function at $M(z_{\text{form}})/M_0 = 0.5$, and is therefore not shown. For haloes with $M_0 < 10^{13} h^{-1} M_{\odot}$, the virial velocity has already reached the present day value when the halo has only assembled 10%-20% of its final mass. Thus, these systems assemble most of their mass without significant changes to the depth of their potential well. Only for massive haloes with $M_0 \gtrsim 10^{14} h^{-1} M_{\odot}$ is the median of $M(z_{\text{vvir}})/M_0$ larger than 0.5, implying that they have assembled the majority of their present day mass through major (violent) mergers.

If we define major mergers as those with a progenitor mass ratio that is at least 1/3, the distribution of $M(z_{\text{lmm}})/M_0$ is remarkably flat. This implies that some haloes accrete a large amount of mass after their last major

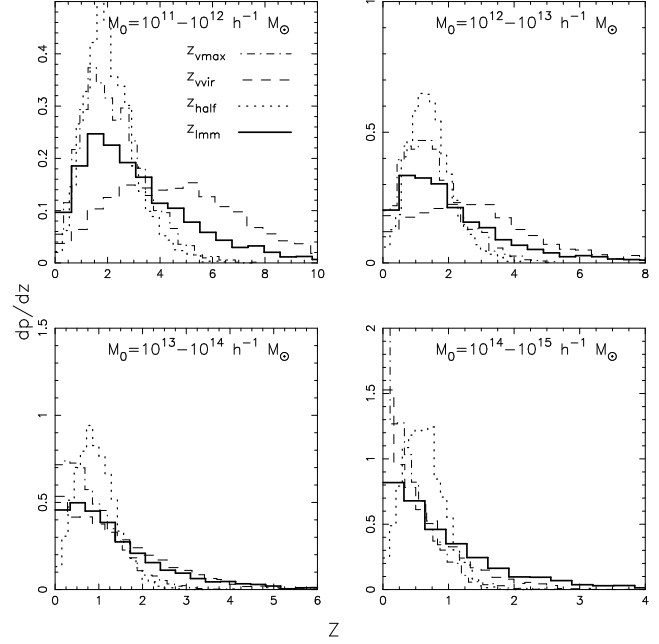


Figure 7. The probability distributions of z_{half} (dotted lines), z_{vvir} (dashed lines), z_{vmax} (dot-dashed lines) and z_{lmm} (thick solid lines). Results are shown for four different mass bins, as indicated in each panel. Note that the scale of the four panels is different! See text for a detailed discussion.

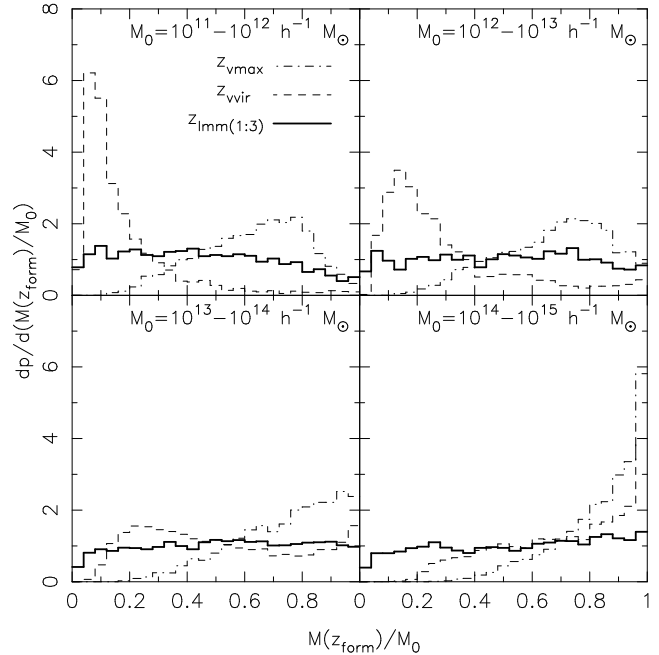


Figure 8. The distributions of the halo mass fraction at various formation times. Different line-styles correspond to different definitions of the formation time, as indicated in the upper left-hand panel. As in Fig. 7, different panels correspond to different halo mass bins, as indicated.

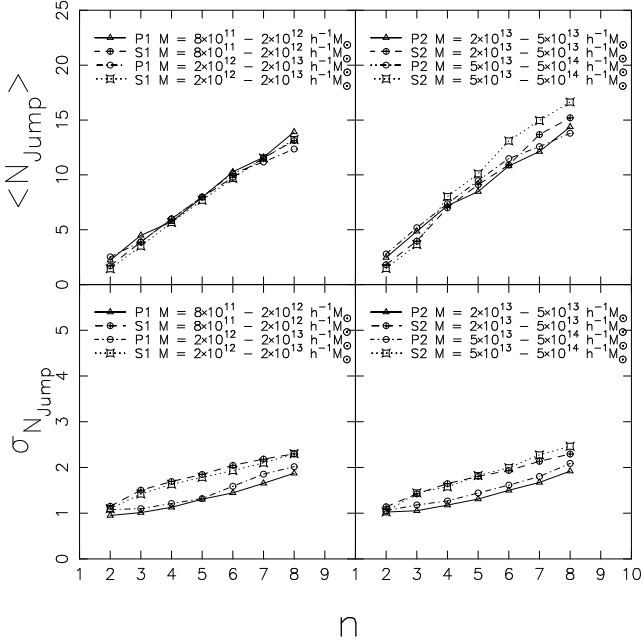


Figure 9. The median, $\langle N_{\text{jump}} \rangle$, and dispersion, $\sigma_{N_{\text{jump}}}$, of the distribution of the number of mass jumps, N_{jump} , in the MAHs, versus n (see text for definitions). Left panels show comparison between P1 and S1, while right panels show comparison between P2 and S2. Note that the agreement between the PINOCCHIO simulations and N -body simulations is remarkable and the mass dependence is rather weak.

merger, while for others the last major merger signals the last significant mass accretion event. Remarkably, the distribution of $M(z_{\text{lim}})/M_0$ is virtually independent of M_0 . For low mass haloes, the flatness of the distribution of $M(z_{\text{lim}})/M_0$ simply reflects the broad distribution of z_{lim} . However, for massive haloes with $M \gtrsim M^*$, the distribution of z_{lim} is fairly narrow. Therefore, for these haloes the flatness of the $M(z_{\text{lim}})/M_0$ distribution implies that, since their last major merger, they have accreted a significant amount of mass due to minor mergers. Since the last major merger occurred fairly recently, this is another indication that massive haloes are still in their fast accretion phase.

5 THE PROPERTIES OF MAJOR MERGERS

During the assembly of dark matter haloes, major mergers play an important role. Not only does a major merger add a significant amount of mass, it also deepens the halo's potential well. Furthermore, in current models of galaxy formation, a major merger of two galaxy-sized haloes is also expected to result in a merger of their central galaxies, probably triggering a starburst and leading to the formation of an elliptical galaxy. Therefore, it is important to quantify the frequency of major mergers during the formation of CDM haloes.

As mentioned above, in a PINOCCHIO simulation mergers of dark matter haloes are treated as instantaneous events, and the masses of the merger progenitors are recorded whenever a merger happens. This makes it very convenient to identify mergers in PINOCCHIO. On

the other hand, in an N -body simulation haloes are identified only in a number of snapshots, and so the accuracy of identifying mergers is limited by the time intervals of the snapshots. For example, if we define major mergers by looking for haloes for which the mass ratio between its second largest and largest progenitors exceeds $1/3$ in the last snapshot, we may miss major mergers in which the two progenitors were assembled during the two snapshots. On the other hand, if we identify major mergers in a simulation by looking for haloes whose masses increase by a factor between $1/4$ and 1 in the next snapshot, we will overestimate the number of major merger events, because some of the haloes may have increased their masses by accretion of small haloes rather than through major mergers. In the simulations used here (S1 and S2), the time intervals between successive snapshots are about 0.3 - 0.6 Gyr, comparable to the time scales of major mergers, and the two definitions of major mergers described above lead to a factor of 2 difference in the number of major mergers. Because of this, it is difficult to make a direct comparison between PINOCCHIO and N -body simulations in their predictions for the number of major mergers. In order to check the reliability of PINOCCHIO in predicting the number of major mergers, we use quantities that are related to the number of major mergers but yet can be obtained from both our N -body and PINOCCHIO simulations. We first construct PINOCCHIO haloes at each of the snapshots of our N -body simulations. We then follow the MAH of each of the present halo using the snapshots and identify the number of events in which the mass of a halo increases by a factor exceeding $1/n$ between two successive snapshots, where n is an integer used to specify the heights of the jumps. In practice, we trace the MAH backward in time until the mass of the halo is 1% of the final halo mass. Since exactly the same analysis can also be carried out for the N -body simulations, we can compare, for a given n and for haloes of given mass at the present time, the statistics of the number of jumps, N_{jump} , predicted by PINOCCHIO simulations with that given by the N -body simulations. We found that the distribution of N_{jump} for a given n can be well fit by a Gaussian distribution, and in Fig. 9 we plot the median $\langle N_{\text{jump}} \rangle$ and standard deviation $\sigma_{N_{\text{jump}}}$ versus n , in several mass bins. The agreement between PINOCCHIO and N -body simulations is remarkably good. Although N_{jump} is not exactly the number of major mergers, the good agreement between PINOCCHIO and N -body simulations makes us believe that it is reliable to use PINOCCHIO to make predictions for the statistics of major mergers.

In order to investigate the statistics of major mergers in detail, we count the number of major mergers for each of the haloes in the ensemble of simulations P0. Here again we only trace a halo back to a time when the mass of its main progenitor is 1% of the halo's final mass. This choice of lower mass limit is quite arbitrary. However, some limit is necessary, because otherwise there will be a large number of major mergers involving progenitors with excessively small masses at very early times. Furthermore this mass limit is also the one we use in defining N_{jump} . The large number of haloes in the ensemble ensures that each mass bin contains about 2000 haloes. Fig. 10 plots the distributions of the number of major mergers (with a progenitor mass ratio $\geq 1/3$) for haloes of different masses at the present time. A halo experiences about 1 to 5 major mergers dur-

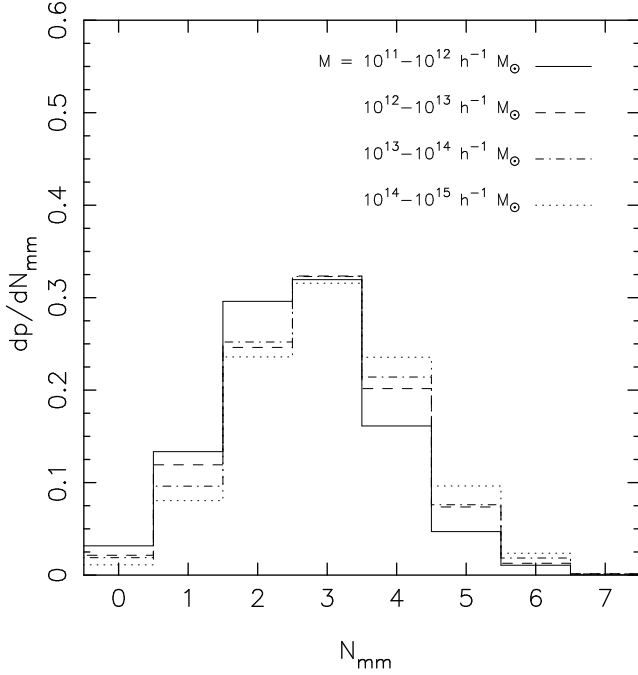


Figure 10. The distribution of the number of major mergers (those with a mass ratio larger than $1/3$) in our PINOCCHIO simulations. Lines in different styles represent different mass bins. Note that the distributions are virtually independent of halo mass.

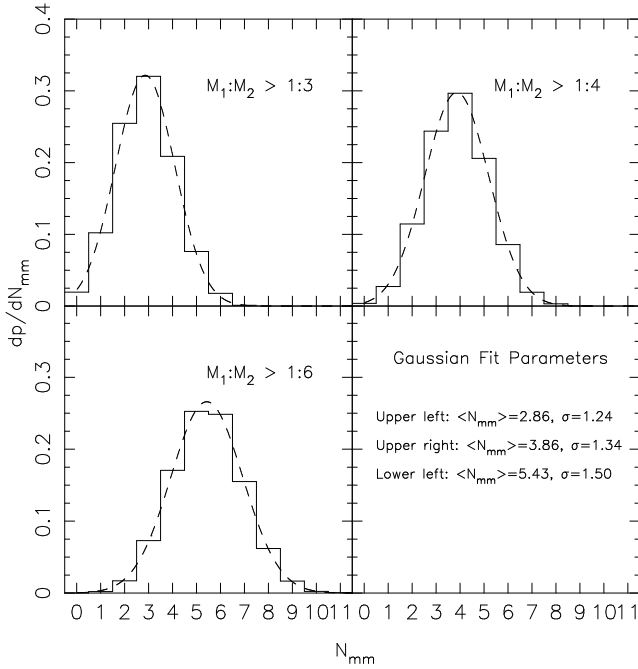


Figure 11. Distribution of the number of mergers (in PINOCCHIO simulations) with a mass ratio larger than $1/3$ (upper left-hand panel), $1/4$ (upper right-hand panel), and $1/6$ (lower left-hand panel). In all three cases all haloes with masses in the range from $10^{11} h^{-1} M_{\odot}$ to $10^{15} h^{-1} M_{\odot}$ are used. The dotted curves show the best-fit Gaussians, the median and standard deviation of which are indicated in the lower right-hand panel.

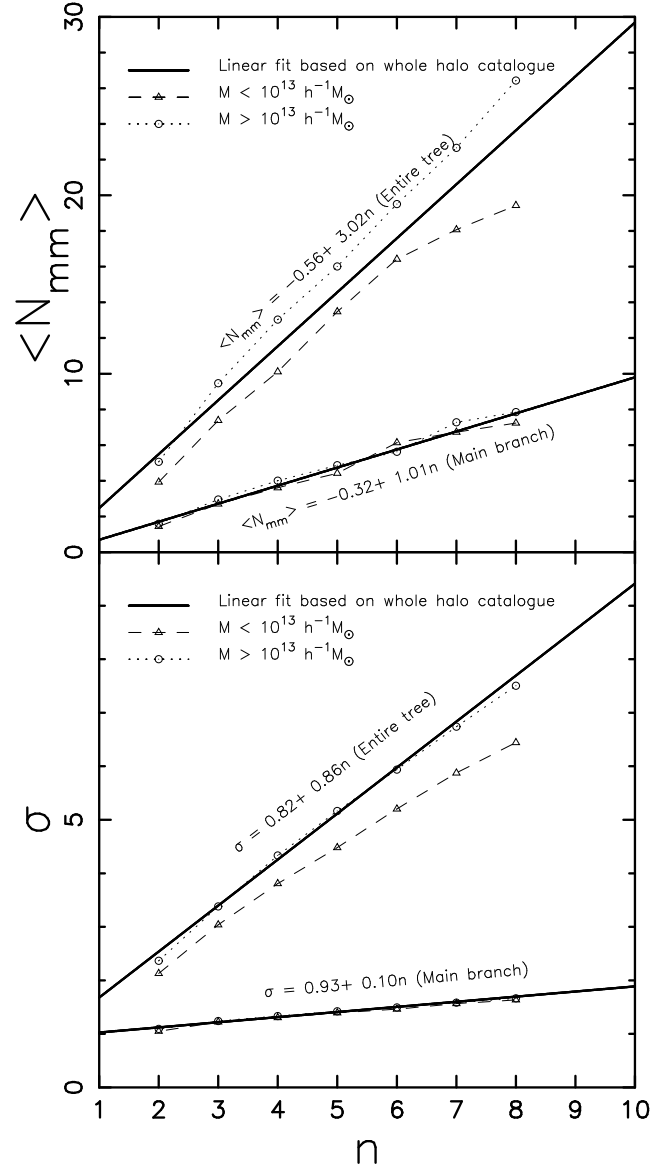


Figure 12. The median (upper panel) and dispersion (lower panel) of the number distributions of mergers with a mass ratio $M_1/M_2 \geq 1/n$, as a function of n . Steeper lines in each panel are the data from all progenitors (summing over all branches of the merger trees) while flatter lines are the results from the main branch. In both cases, we have divided haloes into two mass bins as indicated in each panel. Open triangles connected with dashed lines show the results for haloes with masses $< 10^{13} h^{-1} M_{\odot}$, while open circles connected with dotted lines show the results for haloes with masses $\geq 10^{13} h^{-1} M_{\odot}$. The solid lines are the linear regressions of the data drawn from the whole halo catalogue, with the slopes and zero points indicated.

ing its mass assembly history, with an average of about 3. Note that the N_{mm} -distributions are virtually independent of halo mass. As we have shown in Section 4, however, the redshifts at which these mergers occur do depend strongly on halo mass: while most major mergers occur before $z \simeq 2$ for galaxy-sized haloes, they occur much more recently in the more massive, cluster-sized haloes.

As pointed out above, the progenitor mass ratio used to define a major merger is quite arbitrary. We therefore also

investigate the frequency of mergers with a mass ratio larger than $1/n$ with $n = 2, 4, 5, 6, 7, 8$ (in addition to the $n = 3$ discussed thus far). We find that even with these values of n the distributions of N_{mm} are still virtually independent of halo mass. This allows us to consider a single N_{mm} -distribution for haloes of all masses. Fig. 11 plots these distributions for three different values of n as indicated. Each of these distributions is reasonably well described by a Gaussian function (dashed curves). Note that the use of a Gaussian function is not entirely appropriate, because N_{mm} cannot be negative. However, since the median value of N_{mm} is, in all cases, significantly larger than the width of the distribution, a Gaussian fit is still appropriate. To show how the N_{mm} -distribution depends on n , we plot, as in Fig. 12, the median and the dispersion of this distribution as functions of n . As one can see, both the median and the dispersion increase roughly linearly with n , but the slope for the median (~ 1) is much larger than that for the dispersion (~ 0.1). Note that the results for haloes with masses $< 10^{13} h^{-1} M_{\odot}$ and $> 10^{13} h^{-1} M_{\odot}$ are similar, suggesting the distribution of the number of major mergers is quite independent of halo mass.

Thus far we have only focused on the (major) merger events that merge into the main branch of the merger tree. For comparison, we also consider the merger rates of *all* progenitors, independent of whether they are part of the main branch or not. As before we only consider progenitors with masses in excess of one percent of the final halo mass. The skewer lines in Fig. 12 show the median and dispersion of the number of such mergers as functions of n . Here again, both the median and dispersion have roughly linear relations with n . The median number of such major mergers is roughly three times as high as that of major mergers associated with the main branch, and the dispersion increases with n much faster.

As mentioned above, major mergers are expected to be accompanied by rapid changes of the halo's potential well, due to a resulting phase of violent relaxation. To show such relation in more detail, Fig. 13 shows the distributions of the number of major mergers (defined with $n = 3$) before and after the formation redshift z_{vmax} . For haloes in all mass ranges, only a very small fraction (less than 5%) experiences a major merger at $z < z_{\text{vmax}}$. This demonstrates once again that the growth of the virial velocity is mainly caused by major mergers. This result may have important implications for understanding the structure of dark matter halos. As shown in Lu et al. (2006), if the buildup of the potential well associated with a dark matter halo is through major mergers, then the velocities of dark matter particles may be effectively randomized, a condition that may lead to a density profile close to the universal density profile observed in N -body simulations. Also, if galaxy disks are formed during a period when no major mergers occur, our result suggests that the potential wells of the halos of spiral galaxies should change little during disk formation.

6 CONCLUSIONS

In the current paradigm, galaxies are thought to form in extended cold dark matter haloes. A detailed understanding of galaxy formation, therefore, requires a detailed understanding of how these dark matter haloes assemble. Halo

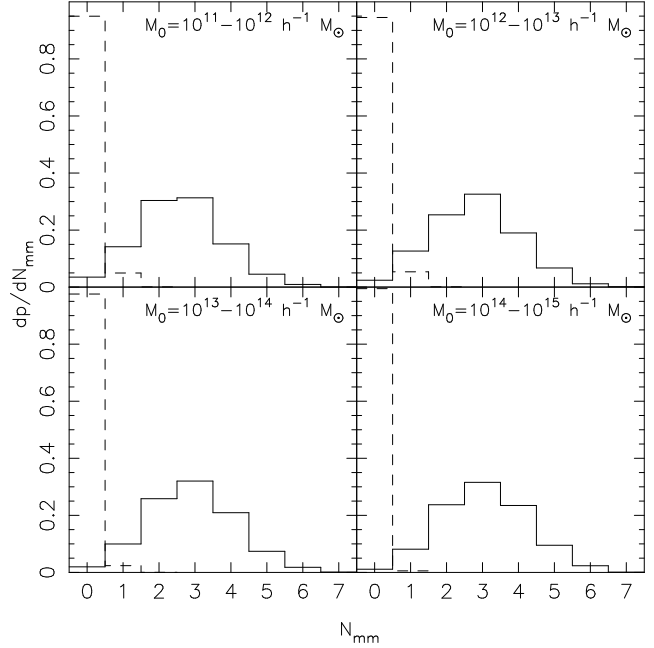


Figure 13. The probability distributions of the number of major mergers (those with a mass ratio larger than $1/3$) before (solid lines) and after (dashed lines) z_{vmax} . Note that the vast majority of major mergers occur at $z > z_{\text{vmax}}$, demonstrating that the growth of the halo's virial velocity is mainly driven by major mergers.

formation histories are typically studied using either numerical simulations, which are time consuming, or using the extended Press Schechter formalism, which has been shown to be of insufficient accuracy. In this paper, we have investigated the growth history of dark matter haloes using the Lagrangian perturbation code PINOCCHIO, developed by Monaco et al. (2002a). We have demonstrated that the mass assembly histories (MAHs) obtained by PINOCCHIO are in good agreement with those obtained using N -body simulations. Since PINOCCHIO is very fast to run, does not require any special hardware such as supercomputers or Beowulf clusters, and does not require any labor intensive analysis, it provides a unique and powerful tool to study the statistics and assembly histories of large samples of dark matter haloes for different cosmologies.

Confirming earlier results based on N -body simulations (e.g. W02; Zhao et al. 2003a,b), we find that typical MAHs can be separated into two phases: an early, fast accretion phase dominated by major mergers, and a late, slow accretion phase during which the mass is mainly accreted from minor mergers. However, the MAHs of individual haloes are complicated, and therefore difficult to parameterize uniquely by a single parameter. We therefore defined four different formation times: the time when a halo acquires half of its final mass, the time when the halo's potential well is established, the time when a halo transits from the fast accretion phase to the slow accretion phase, and the time when a halo experiences its last major merger. Using a large number of MAHs of haloes spanning a wide range in masses, we studied the correlations between these four formation redshifts, as well as their halo mass dependence. Although all four for-

mation times are correlated, each correlation reveals a larger amount of scatter.

For all four formation redshifts, it is found that more massive haloes assemble later, expressing the hierarchical nature of structure formation. Haloes with masses below the characteristic non-linear mass scale, M^* , establish their potential wells well before they have acquired half of their present day mass. The potential wells associated with more massive haloes, however, continue to deepen even at the present time. The time when a halo reaches its maximum virial velocity roughly coincides with the time where the MAH transits from the fast to the slow accretion phase.

If we define major mergers as those with a progenitor mass ratio larger than $1/3$, then on average each halo experiences about 3 major mergers after its main progenitor has acquired one percent of its present day mass. In addition, we found that the number of major mergers the main branch of the merging tree has experienced is linearly correlated with the mass ratio between the merging progenitors. For the whole merging tree, the number of major mergers is about 3 times that of the major mergers in the main branch. The distribution of the number of major mergers a halo has experienced is virtually independent of its mass, and the ratio between the halo mass immediately after the last major merger and the final halo mass has a very broad distribution, implying that the role played by major mergers in building up the final halo can differ significantly from system to system.

ACKNOWLEDGMENTS

We are grateful to Pierluigi Monaco, Tom Theuns and Giuliano Taffoni for making their wonderful code PINOCCHIO publicly available with an easy to understand manual, and to Xi Kang for letting us share his EPS merging tree code. We also thank the Shanghai Supercomputer Center, the grants from NFSC (No. 10533030) and Shanghai Key Projects in Basic Research (No. 05XD14019) for the N-body simulations used in this paper. HJM would like to acknowledge the support of NSF AST-0607535, NASA AISR-126270 and NSF IIS-0611948. FvdB acknowledges useful and lively discussions with Risa Wechsler during an early phase of this project.

REFERENCES

- Alam S.M.K., Bullock J.S., Weinberg D.H., 2002, *ApJ*, 572, 34
- Audit E., Teyssier R., Alimi J.-M., 1997, *A&A*, 325, 439
- Avila-Reese V., Firmani C., 2000, *RevMexAA*, 36, 23
- Bertschinger E., Jain B., 1994, *ApJ*, 431, 486
- Birnboim Y., Dekel A., 2003, *MNRAS*, 345, 349
- Bond J.R., Cole S., Efstathiou G., Kaiser N., 1991, *ApJ*, 379, 440
- Bond J.R., Myers S., 1996a, *ApJS*, 103, 1
- Bond J.R., Myers S., 1996b, *ApJS*, 103, 41
- Borgani S., Coles P., Moscardini L., 1994, *MNRAS*, 271, 223
- Bower R., 1991, *MNRAS*, 248, 332
- Bryan G., Norman M., 1998, *ApJ*, 555, 240
- Buchert T., Ehlers J., 1993, *MNRAS*, 264, 375
- Bullock J.S., Kolatt T.S., Sigad Y., Somerville R.S., Kravtsov A.V., Klypin A.A., Primack J.R., Dekel A., 2001, *MNRAS*, 321, 559
- Catelan P., 1995, *MNRAS*, 276, 115
- Catelan P., Lucchin F., Matarrese S., Porciani C., 1998, *MNRAS*, 297, 692
- Chiueh T., Lee J., 2001, *ApJ*, 555, 83
- Cohn J.D., Bagla J.S., White M., 2002, *MNRAS*, 325, 1053
- Cohn J.D., White M., 2005, preprint (astro-ph/0506213)
- Cole S., Lacey C.G., Baugh C.M., Frenk C.S., 2000, *MNRAS*, 319, 168
- Eisenstein D.J., Loeb A., 1996, *ApJ*, 459, 432
- Eke V.R., Navarro J.F., Steinmetz M., 2001, *ApJ*, 554, 114
- Firmani C., Avila-Reese V., 2000, *MNRAS*, 315, 457
- Governato F., Babul A., Quinn T., Tozzi P., Baugh C.M., Katz N., Lake G., 1999, *MNRAS*, 307, 949
- Gross M.A.K., Somerville R.S., Primack J.R., Holtzman J., Klypin A.A., 1998, *MNRAS*, 301, 81
- Jain B., Bertschinger E., 1994, *ApJ*, 431, 495
- Jenkins A., Frenk C.S., White S.D.M., Colberg J.M., Cole S., Evrard A.E., Couchman H.M.P., Yoshida N., 2001, *MNRAS*, 321, 372
- Kauffmann G., White S.D.M., 1993, *MNRAS*, 261, 921
- Kauffmann G., White S.D.M., Guiderdoni B., 1993, *MNRAS*, 264, 201
- Keres D., Katz N., Weinberg D.H., Dave R., 2005, *MNRAS*, 363, 2
- Lacey C., Cole S., 1993, *MNRAS*, 262, 627
- Lanzoni B., Mamon G.A., Guiderdoni B., 2000, *MNRAS*, 312, 781
- Lee J., Shandarin S., 1998, *ApJ*, 500, 14
- Lin L., Chiueh T., Lee J., 2002, *ApJ*, 574, 527
- Lin W.P., Jing Y.P., Lin L., 2003, *MNRAS*, 344, 1327
- Lu Y., Mo H.J., Katz N., Weinberg M.D., 2006, *MNRAS*, 368, 1931
- Mo H.J., Jing Y.P., White S.D.M., 1996, *MNRAS*, 282, 1096
- Mo H.J., Mao S.D., 2000, *MNRAS*, 318, 163
- Mo H.J., White S.D.M., 1996, *MNRAS*, 282, 347
- Mo H.J., White S.D.M., 2002, *MNRAS*, 336, 112
- Mo H.J., Jing Y.P., White S.D.M., 1997, *MNRAS*, 284, 189
- Monaco P., 1995, *ApJ*, 447, 23
- Monaco P., 1998, *Fundamentals of Cosmic Physics*, Vol. 19, 157
- Monaco P., Theuns T., Taffoni G., Governato F., Quinn T., Stadel J., 2002a, *ApJ*, 564, 8
- Monaco P., Theuns T., Taffoni G., 2002b, *MNRAS*, 331, 587
- Navarro J.F., Frenk C.S., White S.D.M., 1997, *ApJ*, 490, 493
- Nusser A., Sheth R.K., 1999, *MNRAS*, 303, 685
- Peacock J.A., Heavens A.F., 1990, *MNRAS*, 243, 133
- Press W., Schechter P., 1974, *ApJ*, 187, 425
- Rodrigues D.D.C., Thomas P.A., 1996, *MNRAS*, 282, 631
- Sheth R.K., 1998, *MNRAS*, 300, 1057
- Sheth R.K., Lemson G., 1999, *MNRAS*, 305, 946
- Sheth R.K., Mo H.J., Tormen G., 2001, *MNRAS*, 323, 1 (SMT01)
- Sheth R.K., Tormen G., 1999, *MNRAS*, 308, 119
- Sheth R.K., Tormen G., 2002, *MNRAS*, 329, 61

- Somerville R.S., Kolatt T.S., 1999, MNRAS, 305, 1
Somerville R.S., Primack J.R., 1999, MNRAS, 310, 1087
Somerville R.S., Lemson G., Kolatt T.S., Dekel A., 2000, MNRAS, 316, 479
Taffoni G., Monaco P., Theuns T., 2002, MNRAS, 333, 623
Tormen G., 1998, MNRAS, 297, 648
van den Bosch F.C., 2001, MNRAS, 327, 1334
van den Bosch F.C., 2002a, MNRAS, 331, 98
van den Bosch F.C., 2002b, MNRAS, 332, 456
van den Bosch F.C., Mo H.J., Yang X., 2003, MNRAS, 345, 923
Wechsler R.H., Bullock J.S., Primack J.R., Kravtsov A.V., Dekel A., 2002, ApJ, 568, 52 (W02)
White S.D.M., Rees M.J., 1978, MNRAS, 183, 341
Zel'Dovich Y.B., 1970, A&A, 5, 84
Zentner A.R., Bullock J.S., 2002, Phys. Rev. D., 66, 043003
Zhao D.H., Mo H.J., Jing Y.P., Börner G., 2003a, MNRAS, 339, 12
Zhao D.H., Jing Y.P., Mo H.J., Börner G., 2003b, ApJ, 597, 9